Forecasting Tourist Visits During The Covid-19 Pandemic and MotoGP Events Using The Sarima Method

Siti Soraya[1]*, Phyta Rahima[2], Gilang Primajati[3], Maulida Nurhidayati[4], Mohammad Fajri[5]

[1]Computer Science. Faculty of Engineering and Design, Universitas Bumigora
[2]Information System. Faculty of Engineering and Design, Universitas Bumigora
[3]Application Software Engineering. Faculty of Engineering and Design, Universitas Bumigora
[4]Zakat and Waqf Management. Faculty of Islamic Economics and Business, Institut Agama Islam Negeri Ponorogo
[5]Statistics. Faculty of Mathematics and Natural Sciences, Universitas Tadulako
*Corresponding author: sitisorayaburhan@universitasbumigora.ac.id

Received: 27 Februari 2024
Revised: 19 June 2024
Accepted: 23 September 2024

ABSTRACT – **The 5.0 era has made the tourism sector one of the measures of the economic welfare of a region, such as in West Nusa Tenggara (NTB). This is proven by the presence of various types of MSMEs and their innovations and the increasing number of tourist visits to NTB from year to year. The condition of the tourism sector certainly has a positive impact on increasing NTB's economic growth and indirectly on optimizing existing infrastructure. However, extraordinary events such as the earthquake in 2018 and the COVID-19 pandemic resulted in the decline of NTB tourism visits. Then tourist visits in NTB increased again with the holding of the MotoGP Event. The purpose of this study is to forecast the number of tourist visits to NTB. This is very much needed in helping the government to prepare appropriate policies if there is a possibility of a surge in tourist visits in the following years. As well as anticipating if there are other extraordinary events such as earthquakes or global cases. The method used in this study is the Seasonal Autoregressive Integrated Moving Average (SARIMA) Method. The stages in this method are describing data, preprocessing data, identifying stationary models, estimating models, selecting the best SARIMA model, and forecasting with the obtained model to forecast the next desired period. The results of research that have been conducted state that from 2023 to 2024 the number of tourists visiting NTB continues to increase both domestically and abroad. It is hoped that the results of this research will be able to provide information and contribute knowledge and consideration materials in policy making in the development of NTB government tourism.**

Keywords– Tourism, Forecasting, Seasonal Autoregressive Integrated Moving Average (SARIMA), NTB, MotoGP, Covid-19

I. INTRODUCTION

The development of the tourism sector in the 5.0 era is very rapid. Its development is inseparable from the influence of globalization. The development of globalization with various aspects in it also affects the dynamics of the global tourism industry. Related to this, tourism is considered a very profitable sector in improving the economy of a country, because it can encourage economic activity at the local, regional, national, and international levels [1].

Lombok Island is one of the super-priority tourist destinations by being designated as one part of the Special Economic Zone (SEZ) by the Indonesian government. The development of tourism in Lombok is also the axis of tourism development in West Nusa Tenggara Province (NTB). The number of foreign tourists which continues to increase from year to year is inseparable from the rich tourist destinations owned by the island of Lombok itself. In 2013 the number of foreign tourists visiting NTB reached 565,944 million people and reached its peak in 2017 with 1,430,249 million people. Not only foreign exchange income obtained from foreign tourists, but an increase also occurred in local tourist visits. In 2013 local tourists visiting Lombok, NTB reached 791,658 people and in 2017 increased to 2,078,654 people [2].

Lombok's tourism sector in 2018, has a total of 2.8 million visitors which includes 1.6 million local tourists and 1.2 million foreign tourists. The number of visits in 2018 decreased sharply compared to the previous year. With growth of -23% for local tourists and also -16% for foreign tourists. The main factor of the decline in the number of tourist visits, both local and foreign, to Lombok itself was caused by the earthquake that devastated most of Lombok Island during 2018 and COVID-19 cases which indirectly silenced the community's economy in 2019 [3]-[4].

The impact of Covid-19 on Mandalika tourism destinations can trigger a decline in the tourist market. Various policies carried out by tourism institutions to prevent and recover after the pandemic aim to maintain the existence of tourism destinations during and after the Covid-19 pandemic. The development of Mandalika tourism destinations continues to run during the Covid-19 pandemic, especially in the construction of the Mandalika MotoGP Circuit. The MotoGP event at the Mandalika Circuit is believed to be able to bring abundant benefits to Indonesia, especially NTB. By hosting the 2022 MotoGP championship, Indonesia will certainly get enormous branding potential, not only that, this event will also certainly be a promotion of Indonesian tourism as a whole [5]. From this, it can illustrate the importance of forecasting tourist visits to NTB for effective policy planning [6].

The method used to perform forecasting has been used in several studies such as the Seasonal Autoregressive Integrated Moving Average (SARIMA). SARIMA is one of the Box-Jenkins methods by uses observational data behavior and seasonal factors in the data. Reference [7] using the SARIMA method to forecast air quality in Malaysia. Then the same method [8] conducts research related to the government's strategy for overcoming poverty in West Java. Furthermore, Reference [9] also uses the SARIMA approach in forecasting the number of ship passengers at Semayang Port, Balikpapan. Not only here, but SARIMA is also used in forecasting rainfall in Maluku [10]. In forecasting the number of tourist visits in Indonesia, the SARIMa method is used [11].

The results of research that have been conducted state that from 2023 to 2024 the number of tourists visiting NTB continues to increase both domestically and abroad. It is hoped that the results of this research will be able to provide information and contribute knowledge and consideration materials in policy making in the development of NTB government tourism.

The purpose of this study is to forecast the number of tourist visits to NTB from January 2023 to December 2024 using the Sarima Method. The existence of extraordinary events, namely the Covid-19 Pandemic at the end of 2019 until 2020 and the MotoGP Event, a forecast of the number of tourist visits to NTB from January 2023 to December 2024 was carried out using the SARIMA Method. Research related to forecasting is carried out because NTB is one of the priority tourist destinations in Indonesia, which in the end will certainly affect the economy and living order of the people of NTB.

II. LITERATURE REVIEW

A. Box-Jenkins Method

The Box-Jenkins method is a method of applying time series data popularized by George-Box and Gwilyn in the 1970s, by combining moving average and autoregressive approaches [12]. Box - Jenkins recommends that whether or not a time series data is stationary, in the process can later be differentiated 1 or more differentiation processes with the ARIMA model approach [13]. ARIMA is a method that can solve various problems in forecasting time series data, including in forecasting the number of tourists visiting NTB, by looking at past visiting patterns [14].

B. ARIMA Model

The ARIMA model is formed from 3 models, namely Autoregressive (AR), Moving Average (MA), and Autoregressive and Moving Average (ARMA) which is preceded by checking stationary data [15].

If there is an order d ($d \geq 1$) on a forecasting process Z_t hence the homogeneous non-stationary model, known as the ARIMA Model (p, d, q) (16). In general, the ARIMA model can be seen in the equation (1) [16]:

$$\emptyset_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)_{a_t} \quad (1)$$

where AR is represented by:

$$\emptyset_p(B) = (1 - \emptyset_1 B - \cdots - \emptyset_p B^p) \quad (2)$$

and MA is represented by:

$$\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q) \quad (3)$$

In the next section, for the parameters θ_0 show different roles between $d = 0$ and $d \geq 0$. When $d = 0$ contains the meaning that the data is stationary, such as: $\theta_0 = \mu\,(1 - \theta_1 - \cdots - \theta_q)$. However, if $d \geq 0$ known as deterministic trends. A thing that is very rarely used and even eliminated in the forecasting process.

C. Test Independence

Plotting data and identifying the formation of the ARIMA Model is necessary [17]. This is related to stationary in the data is a requirement of the ARIMA Model, both stationary in mean and variant [18]. If the data is not stationary, it is necessary to transform [19]. When the stationary requirements have been met, the identification of the model formed by looking at the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots also needs to be done [20]. Next, the best model estimation stage is carried out which aims to ensure that the residual is white noise or not. The estimation method used is Ljung-Box. It is followed by checking that the model has met the requirements of independent residuals and is normally distributed [16]. Here's the equation for the independence test between residual lags [21]:

$$Q^* = n(n+2)\sum_{k=1}^{K}\frac{\hat{\rho}_k^{\,2}}{(n-k)} \quad (4)$$

with hypotheses:

$H_0\ :\ \rho_1 = \rho_2 = \cdots = \rho_k = 0$ (residual white noise)

H_1 : There is at least one $\rho_k\ \neq 0$, for $k = 1,2,3,\ldots,n$ (residual not white noise)

Equation (4) explains that at the alpha significance level, it is 5%. If the Q * value is greater than the table value $\chi^2{}_{[q;K-p-q]}$ or p-value < alpha, so the decision was made to fail to accept H_0, which means that the residual is not whine noise. In forecasting the properties of future data, this process is carried out by dividing data into in-sample data and out-sample data. Out-sample data is used to forecast or validate the accuracy of data in a natural process. So the model that appears is the best model from in-sample data [22]. In this study, the best model selection measure used SSE and MSE.

D. Seasonal ARIMA Model

Seasonal Additive Model

The seasonal additive model describes seasonal components interacting with non-seasonal components in the model in an additive manner, expressed as SARIMA $\left((\mathrm{p},\mathrm{P}),(\mathrm{d},\mathrm{D}),(\mathrm{q},\mathrm{Q})\right)^s$ for process $\mathrm{Z_t}$ is:

$$(1 - a_1 B - \cdots a_p B^p)(1-B)^d(1-B_s)^D x_t =$$
$$(1 + b_1 B + \cdots + b_q B^q + \theta_s B^s + \cdots + \theta_Q B^{sQ})\varepsilon_t \quad (5)$$

with

B : backward operator $(B^j Y)_t = Y_{t-j}$

p : autoregressive non-seasonal order

q : Non-seasonal order moving average

P : Order of autoregressive coefficients of multiplicative seasonal components

D : Seasonal Differential Order

Q: Order coefficient of moving average of the multiplicative seasonal component

In equation (5), the seasonal component is described by the s-order moving average component. In addition, the seasonal component can also be expressed with an autoregressive component, that is, using a model on the equation (6):

$$\left(1 - a_1 B - \cdots a_p B^p - \beta_p B^{sP}\right)(1 - B)^d (1 - B_s)^D x_t =$$
$$(1 + b_1 B + \cdots + b_q B^q)\varepsilon_t \quad (6)$$

Seasonal Multiplicative Model

In this model, the seasonal component interacts with the non-seasonal component in the model in a multiplicative (multiplication form), as in equation (7):

$$\left(1 - a_1 B - \cdots a_p B^p\right)(1 - B)^d (1 - B_s)^D x_t =$$
$$(1 + b_1 B + \cdots + b_q B^q (1 + \theta_s B^s + \cdots + \theta_Q B^{sQ})\varepsilon_t \quad (7)$$

Here the seasonal component is depicted using the s-order moving average component. Seasonal components can also be expressed using autoregressive, as in the equation (8):

$$\left(1 - a_1 B - \cdots a_p B^p\right)(1 - B)^d (1 - B_s)^D x_t =$$
$$(1 + b_1 B + \cdots + b_q B^q (1 + \theta_s B^s + \cdots + \theta_Q B^{sQ})\varepsilon_t \quad (8)$$

This multiplicative model can be denoted as an ARIMA model (p,d,q) $(\mathrm{P}, \mathrm{D}, \mathrm{Q})^s$.

III. METHODOLOGY

In this study, the SARIMA Method was used. which serves to predict the number of tourist visits to West Nusa Tenggara Province (NTB). The data used as predictions is monthly data on tourist visits from January 2012 to December 2022. In the process, the data is divided into 2 parts, namely data in sample and data out sample. Data in sample starts from January 2012 to December 2022, while data out sample starts from January 2022 to December 2022. The data in this study was sourced from the Central Bureau of Statistics of West Nusa Tenggara through the https://ntb.bps.go.id/ page. The stages of research carried out are as follows [22]:

a. Describe the available data.

b. Preprocessing data.

c. Identify stationary models.

d. Perform model estimation.

e. Selecting the best SARIMA model

f. Perform forecasting with the obtained model.

g. Perform forecasting of the next desired period.

IV. RESULTS AND DISCUSSION

This study used data on the number of tourist visits to West Nusa Tenggara (NTB) from January 2012 to December 2022. This data was obtained from the NTB Tourism Office. A description of the number of tourist visits is shown in Figure 1.

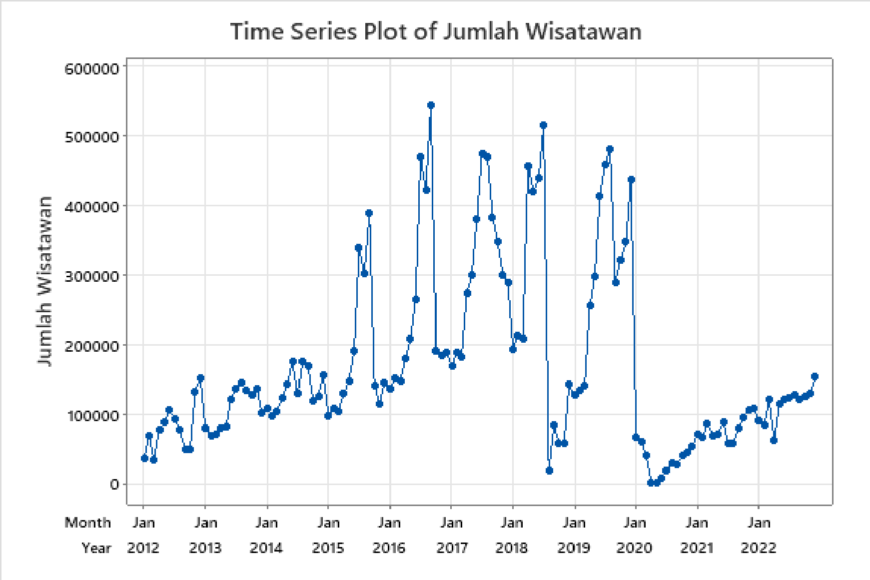

Figure 1 Plot of Number of Tourist Visits to NTB

The development of the number of tourist visits to NTB has been presented in Figure 1. This provides information that the number of tourist visits has a seasonal trend, which tends to increase in April and August and often decreases in February and November. Based on tourist visit data for the last 3 years, it is known that the month. December 2019 showed the highest number of visits compared to other previous months. Then it dropped drastically during 2020 due to the Covid-19 Pandemic hitting almost all regions in the world. The decline in the number of tourism visits did not last long, because the government revived tourism with the holding of MotoGP events in 2022. Based on the description of Figure 1, the Seasonal ARIMA analysis was carried out.

The first thing to do is to test the stationarity of data in both the mean and the variant. Stationarity testing in variants is performed by Box-Cox testing, as presented in Figure 2.

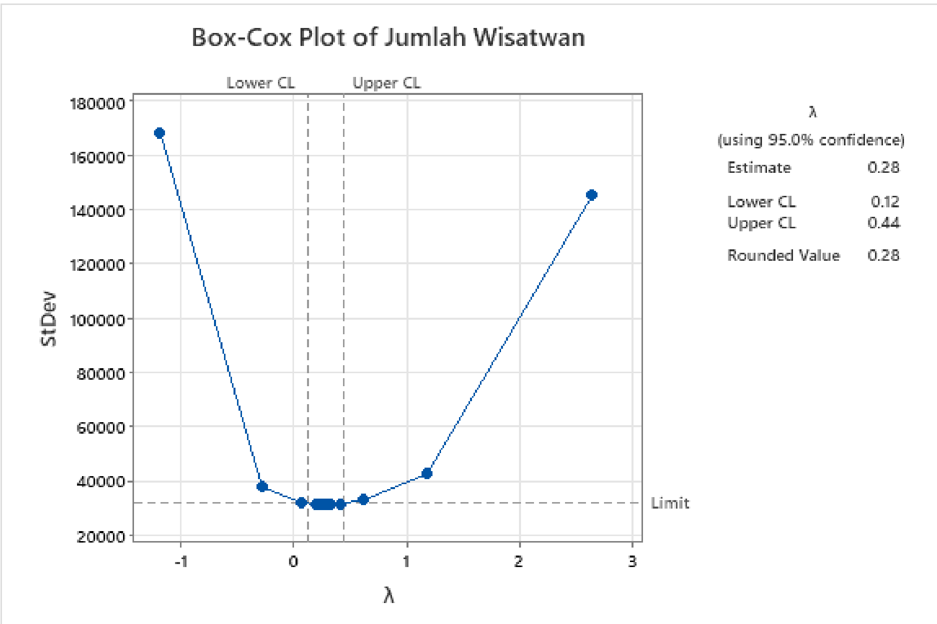

Figure 2. Box-Cox Test Results

Figure 2 shows that the lambda value is 0.28 which means that the transformation must be performed with the logarithmic function. The transformation data is then tested for stationarity in mean using ACF and PACF as shown in Figure 3.

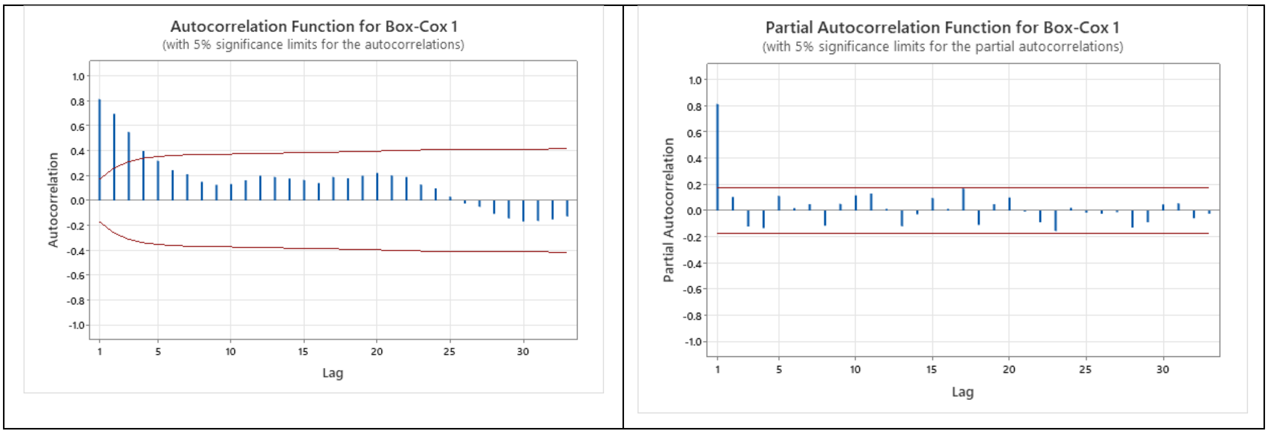

Figure 3. Plot ACF and PACF Data Transformations

The results of the ACF and PACF plots in Figure 3 show that the ACF drops rapidly and the PACF cuts off at lag 1 so that the transformation data is not stationary in the mean so it is necessary to differentiate the data. The results of ACF and PACF plot differencing data are shown in Figure 4.

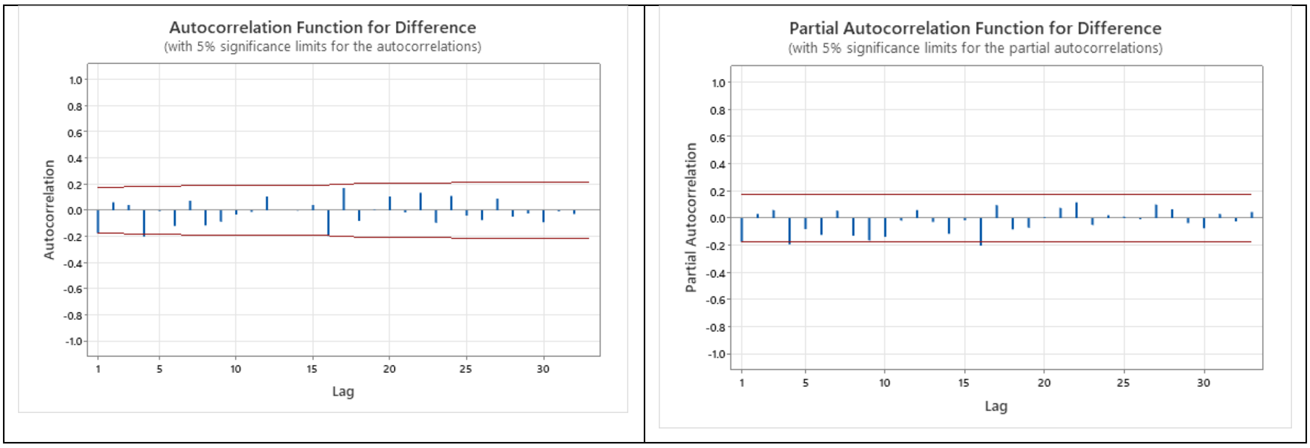

Figure 4. Plot ACF and PACF Data Differencing

Figure 4 shows that the ACF and PACF plots experienced cuts of lag 4 and 16 so that the data on the number of tourist visits to NTB followed the SARIMA model (0,1,0) (2, 0, 2)4. Then proceed to estimate several models formed as in Table 1

Table 1. ARIMA Seasonal Model (SARIMA) Estimation Results

Model	Type	Coef	P-Value
SARIMA(0,1,0)(0,0,2)^4		0.219	0.014
	SMA (8)	0.145	0.103
SARIMA(0,1,0)(1,0,2)^4	SAR (4)	-0.761	0.000
	SMA (4)	-0.592	0.000
	SMA (8)	0.334	0.000
SARIMA(0,1,0)(2,0,0)^4	SMA (4)	-0.223	0.012
	SMA (8)	-0.143	0.107
SARIMA(0,1,0)(2,0,1)^4	SAR (4)	0.383	0.300
	SMA (4)	-0.010	0.945
	SMA (8)	0.616	0.088
SARIMA(0,1,0)(2,0,2)^4	SAR (4)	-0.632	0.017
	SAR (8)	0.117	0.660
	SMA (4)	-0.464	0.053
	SMA (8)	0.459	0.068

Table 1 describes several estimated models, namely: SARIMA model (0,1,0) (0,0,2)4 with 1 significant parameter, SARIMA model (0,1,0) (1,0,2)4 with all significant parameters, SARIMA model (0,1,0) (2,0,0)4 with 1 parameter, SARIMA(0,1,0) model (2,0,1)4 only 1 significant parameter, finally SARIMA(0,1,0) (2,0,2)4 model with 2 significant parameters. Then with diagnostic check to ensure the model obtained from the estimation results meets the assumption of white noise using Ljung-Box testing.

Table 2. White Noise Test Results

Model	Lag	Chi-Square	P-Value
SARIMA(0,1,0)(0,0,2)^4	12	50.12	0.000
	24	81.36	0.000
	36	88.40	0.000
	48	107.22	0.000
SARIMA(0,1,0)(1,0,2)^4	12	50.29	0.000
	24	78.24	0.000
	36	83.32	0.000
	48	100.25	0.000
SARIMA(0,1,0)(2,0,0)^4	12	51.75	0.000
	24	85.90	0.000
	36	91.58	0.000
	48	108.01	0.000
SARIMA(0,1,0)(2,0,1)^4	12	51.08	0.000
	24	83.04	0.000
	36	91.42	0.000
	48	109.34	0.000
SARIMA(0,1,0)(2,0,2)^4	12	48.76	0.000
	24	75.52	0.000
	36	81.13	0.000
	48	99.31	0.000

Based on Table 2 it is known that all Q statistics for lag 12, 24, 36, and 48 for all SARIMA models are significant so that all SARIMA models are white noise. To determine the best model of these SARIMA models, it is necessary to test the performance of the model. The smaller the AIC value indicates that it is the best model to use. The results of the SARIMA model performance test are shown in Table 3.

Table 3. SARIMA Model Performance Test Results

Model	AIC
SARIMA(0,1,0)(0,0,2)^4	3.301
SARIMA(0,1,0)(1,0,2)^4	3.238
SARIMA(0,1,0)(2,0,0)^4	3.306
SARIMA(0,1,0)(2,0,1)^4	3.302
SARIMA(0,1,0)(2,0,2)^4	3.237

Based on Table 3 it is known that the AIC value of the SARIMA model (0,1,0)(2,0,2)4 is smaller than other SARIMA models, so the best SARIMA model used for further analysis is the SARIMA model (0,1,0)(2,0,2)4. Figure 4 is a plot of SARIMA model forecasting data (0,1,0)(2,0,2)4.

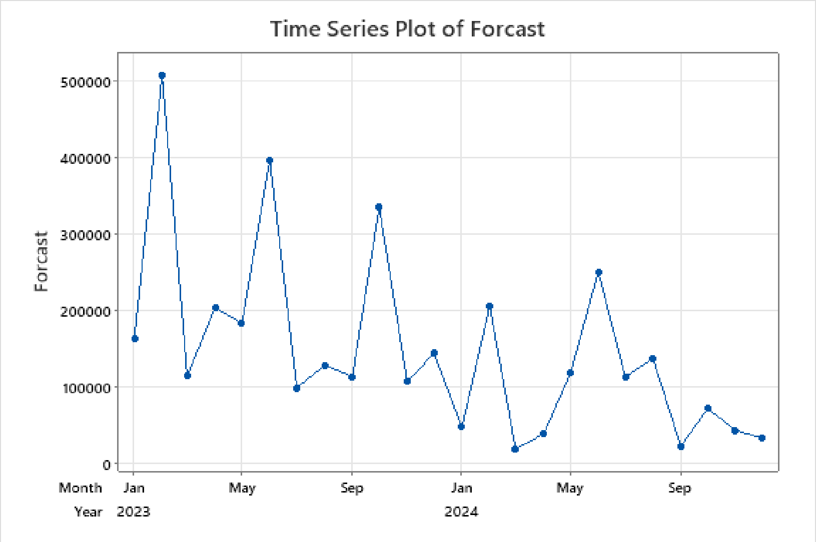

Figure 5. Forecast Data Plot 2023-2024

The results of predictions made with the SARIMA model in Figure 5 show that the SARIMA model can follow the pattern of the data it has. In this case, it is data on the transformation of the number of tourist visits. After obtaining the optimal SARIMA model, then forecasting for the next 24 months and obtaining an AIC value of. This performance value is higher than the performance value of the data in the sample. At this stage, it is continued by displaying the forecasting results for the number of tourist visits to NTB from January 2023 to December 2024 with the SARIMA model (0,1,0)(2,0,2)4. The forecasting results are shown in Table 4.

Table 4. Forecasting the Number of Tourism Visits with the SARIMA Model (0,1,0) (2, 0, 2)^4

2023	Value	2024	Value
January	162878	January	47632
February	506537	February	205017
March	114587	March	19388
April	204013	April	38626
May	183150	May	117992
June	397025	June	250479
July	99010	July	113701
August	128539	August	137273
September	114029	September	23006
October	334688	October	72014
November	107570	November	43347
December	143860	December	33354

Table 4 shows the forecast for the number of tourist visits to NTB in 2023 to increase the most in February with the number of visits of 506537 people and the smallest number of visits occurred in July of 99010 people. On the other hand, the highest number of tourist visits in 2024 will occur in June as many as 250479 people and the smallest in March as many as 19388 people.

V. CONCLUSION

The conclusion that can be drawn from the results of research that has been carried out is that the number of tourist visits to NTB in 2023 has increased the most in February with the number of visits of 506537 people and the smallest number of visits occurred in July of 99010 people. On the other hand, the highest number of tourist visits in 2024 will occur in June as many as 250479 people, and the smallest in March as many as 19388 people.

VI. SUGGESTION

The suggestion from this study for future research is to try to compare the SARIMAX method with the VARIMAX model by involving regional aspects.

REFERENCES

1. Valguna PA, Dewanti D, Suparma LB. Dampak Perkembangan Pariwisata Pulau Lombok terhadap Pengembangan Bandar Udara Internasional Lombok. J Manaj Aset Infrastruktur Fasilitas. 2020;4(3):195–210.
2. Moerwanto AS, Junoasmono T. Strategi Pembangunan Infrastruktur Wisata Terintegrasi. J HPJI. 2017;3(2):67–78.
3. Irfan P, Apriani A. Analisa Strategi Pengembangan E-Tourism Sebagai Promosi Pariwisata Di Pulau Lombok. Ilk J Ilm. 2017;9(3):325–30.
4. Kadarisman A. Government public relations dalam pengembangan pariwisata masa pandemi COVID-19 di Geopark Ciletuh. PRofesi Humas J Ilm Ilmu Hub Masy. 2021;5(2):270.
5. Hayatun Sofian, Yulanda Trisula SY, Sandi Justitia Putra, Lalu Ahmad Rahmat, I Made Putra Suryantara. Analisis Agenda Setting Pemberitaan MotoGP Mandalika Pada RRI Mataram. MUKASI J Ilmu Komun. 2022;1(1):11–20.
6. Herawati S. Peramalan Kunjungan Wisatawan Mancanegara Menggunakan Generalized Regression Neural Networks. J InfoteL - Inform Telekomun Elektron. 2016;8(1):35.
7. Rahman NHA, Lee MH, Suhartono, Latif MT. Evaluation Performance of Time Series Approach for Forecasting Air Pollution Index in Johor, Malaysia. Sains Malaysiana. 2016;45(11):1625–33.
8. Silalahi DK. Forecasting of Poverty Data Using Seasonal ARIMA Modeling in West Java Province. JTAM | J Teor dan Apl Mat. 2020;4(1):76.
9. Multiningsih M, Siswanah E, Saleh M. Forecasting the Number of Ship Passengers with SARIMA Approach (A Case Study: Semayang Port, Balikpapan City). JTAM (Jurnal Teor dan Apl Mat. 2022;6(4):1060.
10. Kasus S, Hujan C, Ambon K, Maluku P, Kafara Z, Rumlawang FY, et al. Peramalan Curah Hujan Dengan Pendekatan Seasonal Autoregressive Integrated Moving Average (Sarima) Rainfall Forecasting Using Seasonal Autoregressive Integrated Moving Average Model (Sarima). 2017;11:63–74.
11. Ilmayasinta N. Forecasting Arrival of Foreign Tourists Using Seasonal Arima Box-Jenkins. BAREKENG J Ilmu Mat dan Terap. 2021;15(2):223–30.
12. City S, City S, City S, Tahir HA, Ahmed RA, Mhamad AJ. Forecasting for the Imported Weight of Equipment to Cargo of Sulaimani International Airport. 2019;2520(3):62–82.
13. Eni D, Adeyeye FJ. Seasonal ARIMA Modeling and Forecasting of Rainfall in Warri Town, Nigeria. J Geosci Environ Prot. 2015;03(06):91–8.
14. Didiharyono D, Bakhtiar B. Forecasting Model With Box-Jenkins Method To Predict Tourists Who Visit Tourism Place In Toraja. J Econ Manag Account. 2018;1(1):34–41.
15. Chuang A, Wei WWS. Time Series Analysis: Univariate and Multivariate Methods. Vol. 33, Technometrics. 1991. p. 108.
16. Annisa Fitri, Ika Purnamasari MS. Peramalan jumlah wisatawan mancanegara menggunakan model arima. Statistika. 2019;7(1).
17. Zhang Y, Luo L, Yang J, Liu D, Kong R, Feng Y. A hybrid ARIMA-SVR approach for forecasting emergency patient flow. J Ambient Intell Humaniz Comput. 2019;10(8):3315–23.
18. Arini PS, Nawangsih E. Peramalan Jumlah Kunjungan Wisatawan Mancanegara (Wisman) ke Bali Tahun 2019: Metode ARIMA. 2019;136–41.
19. Didiharyono D, Syukri M. Forecasting with arima model in anticipating open unemployment rates in south sulawesi. Int J Sci Technol Res. 2020;9(3):3838–41.
20. Li W, Zhang ZG. Based on time sequence of ARIMA model in the application of short-term electricity load forecasting. ICRCCS 2009 - 2009 Int Conf Res Challenges Comput Sci. 2009;11–4.
21. Iriani NP, Akbar MS. Estimasi Value at Risk (VaR) pada Portofolio Saham dengan Copula. 2010;(3):1–6.

22. Hendayanti NPN, Nurhidayati M. Perbandingan Metode Seasonal Autoregressive Integrated Moving Average (SARIMA) dengan Support Vector Regression (SVR) dalam Memprediksi Jumlah Kunjungan Wisatawan Mancanegara ke Bali. J Varian. 2020;3(2):149–62.